Affine Reduction of Dimensionality: An Origin-Centric Perspective

Robert L. Obenchain, Risk Benefit Statistics LLC, wizbob@att.net

Abstract

We consider statistical methods for reduction of multivariate dimensionality that have invariance and/or commutativity properties under the affine group of transformations (origin translations plus linear combinations of coordinates along initial axes). The methods discussed here differ from traditional principal component and coordinate approaches in that they are origin-centric. Because all Cartesian coordinates of the origin are zero, it is the unique fixed point for subsequent linear transformations of point scatters. Whenever visualizations allow shifting between and/or combining of Cartesian and polar coordinate representations, as in Biplots, the location of this origin is critical. Specifically, origin-centric visualizations enhance the psychology of graphical perception by yielding scatters that can be interpreted as Dyson swarms. The key factor is typically the analyst's choice of origin via an initial "centering" translation; this choice determines whether the recovered scatter will have either no points depicted as being near the origin or else one (or more) points exactly coincident with this origin.

Keywords: Affine Transfomations, Oblique and Orthogonal Projections, Principal Components and Coordinates, Multidimensional Scaling, Euclidean distance, Cartesian and polar coordinates.

Contents

1 Introduction

We discuss an intuitive alternative to classical Principal Components analysis that selects the (singular) affine transformation of the original scatter, visualized as N points embedded within a p-dimensional Euclidean space, that reduces its dimensionality to q < p while attempting to reproduce all pair-wise, squared Mahalanobis inter-point distances. We show that this criterion is not coordinate-free; it places extra emphasis upon reproducing the squared Euclidean distances of all N points from the origin. Our origin-centric perspective on affine reduction of dimensionality encourages interpretation of the recovered scatter using either polar or Cartesian coordinates.

In traditional principal component and coordinate approaches to reduction of dimensionality, the origin is conventionally placed at the geometric centroid of the overall scatter by subtracting the observed arithmetic mean value from each given coordinate. This convention can, optionally, be retained in the origin-centric approach discussed here. On the other hand, the impact of details such as how the given coordinates were originally scaled and intercorrelated are minimized in this new, alternative approach. This is accomplished by placing the given data scatter in its affine invariant canonical form, Obenchain(1971, 1972a). Interestingly, this substitution makes reduction of dimensionality using conventional principal components analysis impossible; all one-dimensional orthogonal projections of the revised data have the same variance, and any two mutually orthogonal projections are uncorrelated. In particular, this means that the ultimate impact of researcher initiatives, Glaeser(2006), in choice of included variables and of their initial scaling upon potential recovered configurations of low dimensionality is thereby greatly reduced.

The format of the paper is as follows. In the next three Sections, we introduce basic notation and define terminology related to Euclidean distance calculations, matrix norms in linear algebra, and origin-centric coordinate concepts. Sections 5 and 6, then display two relatively simple numerical examples. We end this overview paper

with some general discussion points in Section 7. The final four Sections, 8 to 11, are somewhat more technical Appendices covering historical background information and some relatively advanced observations and conjectures.

2 Origin-Centric Coordinates and Euclidean Interpoint Distances

When a N × p matrix of finite real numbers, $\mathbf{X}$, is used as a Cartesian coordinate representation of N points in Euclidean space relative to p = 2 or more mutually orthogonal axes, the conventional view is that the location of the origin is not important. After all, the Pythagorean theorem dictates that the squared distance between the i-th and j-th points in the scatter is given by

$$\begin{aligned} \mathbf{d}_{ij}^2 &= (\mathbf{x}_i' - \mathbf{x}_j')(\mathbf{x}_i - \mathbf{x}_j) \\ &= \mathbf{x}_i'\mathbf{x}_i + \mathbf{x}_j'\mathbf{x}_j - 2\mathbf{x}_i'\mathbf{x}_j, \end{aligned} \tag{2.1}$$

where $\mathbf{x}_i'$ denotes the i-th row of $\mathbf{X}$. In the first line of the above expression, note that only differences in coordinates between points are involved. It follows that $\mathbf{d}_{ij}^2$ is, by definition, invariant under shifts in the location of the origin resulting from simple (additive) translations of coordinates.

The second line of equation (2.1) shows an equivalent expression containing origin-centric terms. The first two terms are the squared Euclidean distances of the i-th and j-th points from their current origin, $\mathbf{0}$.

Similarly, the full N × N matrix of inter-point squared distances can be written as

$$\begin{aligned} \mathbf{D}^{(2)} &= ((d_{ij}^2)) \text{ ...for all i and j from 1 to N} \\ &= (\mathbf{d}_0^2)\mathbf{1}' + \mathbf{1}(\mathbf{d}_0^2)' - 2\mathbf{X}\mathbf{X}', \end{aligned} \tag{2.2}$$

where $\mathbf{d}_0^2$ is the N × 1 column vector that consists of the ordered elements of $Diag(\mathbf{X}\mathbf{X}')$. In other words, the i-th element of $\mathbf{d}_0^2$ is the $\mathbf{x}_i'\mathbf{x}_i$ term of equation (2.1). Thus $\mathbf{d}_0^2$ is the vector of squared Euclidean distances of all N points from their current origin, $\mathbf{0}$. The $\mathbf{D}^{(2)}$ matrix is symmetric, and its N diagonal elements are null.

Suppose now that a given $\mathbf{X}$ matrix is to be optimally approximated by a N × q matrix of finite real numbers, $\mathbf{Z}$, where q < p but q is at least 1. Specifically, suppose that the objective is to minimize the squared Frobenius norm of the difference between

the $\mathbf{D}^{(2)}$ matrices computed from the $\mathbf{X}$ and $\mathbf{Z}$ coordinates:

$$\mathbf{Norm}^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}(d_{ij}^2(\mathbf{X}) - d_{ij}^2(\mathbf{Z}))^2$$
$$= ||\rho\mathbf{1}' + \mathbf{1}\rho' - 2(\mathbf{XX}' - \mathbf{ZZ}')||^2$$
$$= 4||\mathbf{XX}' - \mathbf{ZZ}'||^2 + 2(\rho'\mathbf{1})^2 + 2N\rho'\rho, \tag{2.3}$$

where $\rho = d_0^2(\mathbf{X}) - d_0^2(\mathbf{Z})$ is the column vector of discrepancies in squared difference of points from the origin in $\mathbf{X}$ and $\mathbf{Z}$ coordinates.

In the special case of primary interest here, the $\mathbf{XX}'$ outer-products matrix ($N \times N$) is what is know as an Association matrix, $\mathbf{A}$, for "individuals" (points), Gower(1966), while the $\mathbf{X}'\mathbf{X}$ inner-products matrix ($p \times p$) contains the corresponding Adjusted Sums-of-Squares and Cross-Products for "variables." The objective in Principal Component Analysis (PCA) is to minimize $||\mathbf{A} - \mathbf{ZZ}'||^2$ where the $\mathbf{X}$ and $\mathbf{Z}$ matrices have each been "centered" as outlined below in Section 3. This particular point is more fully discussed in an Appendix; see Section 9.

In any case, minimizing $\mathbf{Norm}^2$ of equation (2.3) is potentially quite different from ordinary PCA. Minimizing $\mathbf{Norm}^2$ places emphasis upon balancing two separate criteria: [1] reducing the size of the ρ vector, which contains the discrepancies in *squared distance from the origin* for all N points within the original $\mathbf{X}$ and derived $\mathbf{Z}$ configurations, and [2] reducing the PCA norm of the difference in outer-product matrices. Tension between these competing objectives make choice of $\mathbf{Z}$ so as to minimize $\mathbf{Norm}^2$ of equation (2.3) an origin-centric approach.

3   Affine Reduction of Multivariate Dimensionality

The *Singular Value Decomposition* of a N $\times$ p matrix consisting of the "centered," real valued $\mathbf{X}$ coordinates of N points in a Euclidean space of dimension (linear rank) r can be written as:

$$(\mathbf{I} - \mathbf{1}\gamma')\mathbf{X} = \mathbf{H}\mathbf{\Lambda}^{1/2}\mathbf{G}', \tag{3.1}$$

where the $1 \times N$ row vector γ' is any generalized inverse of the N $\times$ 1 column vector of all ones, $\mathbf{1}$, the semi-orthogonal matrix $\mathbf{H}$ is the N $\times$ r matrix of standardized *Principal Coordinates*, $\mathbf{\Lambda}^{1/2}$ is a r $\times$ r diagonal matrix of strictly positive singular values, and $\mathbf{G}'$ is a r $\times$ p semi-orthogonal matrix of component *Direction Cosines*. When r = p, the $\mathbf{G}$ matrix becomes orthogonal (and invertible); $\mathbf{GG}' = \mathbf{I}$ of order p.

Every choice for γ is such that $\gamma'\mathbf{1} = 1$; it follows from (3.1) that the semi-orthogonal matrix $\mathbf{H}$ is orthogonal to γ. This $\gamma'\mathbf{H} = 0$ restriction assures that r $\leq min(p, N - 1)$. More complete information about this critical choice of initial translation is given in the Appendix of Section 9; choice of γ determines the location of the origin, $\mathbf{0}$, of Cartesian coordinates.

Affine reductions in dimensionality are those in which the recovered $\mathbf{Z}$ configuration is of the restricted form:

$$\mathbf{Z} = \mathbf{HB}, \tag{3.2}$$

where $\mathbf{H}$ is the N × r matrix of (3.1) and $\mathbf{B}$ is a r × q matrix of $rank(\mathbf{B}) = q < r$ consisting of real scalars. In particular, any row of $\mathbf{H}$ that is null assures that the corresponding row of $\mathbf{Z}$ will also be null ...corresponding to a point coincident with the origin, $\mathbf{0}$.

One of two forms of preliminary data "standardization" are typically applied prior to performing traditional PCA, but neither is nearly as "drastic" as replacing one's initial $\mathbf{X}$ matrix by its Mahalanobis principal coordinates matrix, $\mathbf{H}$, as in (3.2). Traditional "mean centering" uses the $\mathbf{1}^+$ vector, with each of its N elements equal to 1/N, as choice of γ' in (3.1). Traditional standardization to "correlation form" performs an addition step in which each column of the "mean centered" $\mathbf{X}$ matrix is then divided by its sample standard deviation. This is essentially equivalent to the standardization in which the N elements in each column of $\mathbf{X}$ are made to not only sum to zero but also to have sum-of-squares set equal to 1; on the other hand, this standardization still allows the different columns of $\mathbf{X}$ to be (linearly) correlated or confounded.

In sharp contrast, the semi-orthogonal $\mathbf{H}$ matrix of (3.1) is essentially r-dimensional from the perspective of traditional PCA. Each column of this $\mathbf{H}$ matrix is orthogonal to the current choice of γ (defining a mean or multivariate median measure of location), has sum-of-squares set equal to 1, and is exactly uncorrelated with the derived coordinates along any direction strictly orthogonal to it.

For example, consider the 1-dimensional case, q = 1 in (3.2), where the $\mathbf{Z}$ and $\mathbf{B}$ matrices become single column vectors, $\mathbf{z}$ and $\mathbf{b}$. It follows that $\gamma'\mathbf{z} = 0$ while the corresponding sum-of-squares is $\mathbf{z}'\mathbf{z} = \mathbf{b}'\mathbf{H}'\mathbf{Hb} = \mathbf{b}'\mathbf{b}$, which is determined solely by the squared length of the $\mathbf{b}$ vector. Similarly, any two different choices for this $\mathbf{b}$ vector yield uncorrelated $\mathbf{z}$ representations if and only if these two choices are mutually orthogonal vectors, $\mathbf{b_1}'\mathbf{b_2} = 0$.

In other words, since PCA cannot be used to reduce the dimensionality of $\mathbf{H}$, one's primary hope may well be to adopt an origin-centric perspective based upon minimizing the squared norm of equation (2.3).

4   Adding a Point at the Origin

If none of the rows of the $\mathbf{H}$ matrix of (3.1) is null, then the given scatter contains no points coincident with the origin, as determined by the current choice of γ vector. On the other hand, there will always be at least one point at the current origin whenever exactly one element of the current γ vector is a 1 and all other elements are null.

Any null row of the $\mathbf{H}$ matrix of (3.1) can be relocated to its first row simply by renumbering the N points in the initial scatter. The resulting $d_0^2(\mathbf{X})$ vector of (2.3)

that contains the squared Euclidean distances of all N point from their current origin will then naturally have a zero in its first row. In fact, this $d_0^2(\mathbf{X})$ vector will also then be the first column of the $\mathbf{D}^{(2)}$ matrix of (2.2), while its transpose will be the first row of this same matrix.

When a $\mathbf{H}$ matrix from (3.1) contains no null row, a null row can be added as, say, row N+1. This augmented $\mathbf{H}$ matrix is still semi-orthogonal, and $\mathbf{H'H}$ is still the r $\times$ r identify matrix. The corresponding augmented γ column vector $((N+1) \times 1)$ can be taken to have a zero as its final element, and the augmented $\mathbf{D}^{(2)}$ matrix of (2.2) will have the original $d_0^2(\mathbf{X})$ vector as the above the diagonal elements in its last column ...and its transpose in the bottom row.

In summary, the $d_0^2(\mathbf{X})$ vector that contains the squared Euclidean distances of all N original points from their current origin is either already contained within the original N $\times$ N squared distances matrix, $\mathbf{D}^{(2)}$ of (2.2), or else it will appear within the corresponding (N+1) $\times$ (N+1) augmented version of this matrix when a new point coincident with the origin is included.

In other words, the $d_0^2(\mathbf{X})$ vector is either an explicit or implicit part of every N $\times$ N matrix of squared distances, $\mathbf{D}^{(2)}$ of (2.2). In sharp contrast with PCA, efforts to minimize $\mathbf{Norm}^2$ of (2.3) place special emphasis upon accurate reproduction of $d_0^2(\mathbf{X})$ values within the recovered Z-configuration.

5   First Numerical Example: Six Points on a Plane

Perhaps the most simple sort of "interesting" example for comparing PCA with our origin-centric approach is depicted as H coordinates (p=2 and N=6) in Figure 1. When reducing this configuration to a single dimension (q=1), a singular affine transformation can produce either "a pair of triplet points" or else a "triplet of pairs." In fact, these particular configurations happen to correspond to two of the three local minima of $\mathbf{Norm}^2$ of equation (2.3); see Figure 2.

This example is also interesting in the sense that, while the initial H-configuration contains no point(s) at the origin, the optimal q=1 dimensional recovered configuration [case (c) in Figure 2] contains 2 points coincident with the origin.

6   Second Numerical Example: The Longley Data in q=2 Dimensions

Here we use the infamous Longley(1972) dataset (p=6, N=16) to illustrate some key distinctions between the PCA and affine, origin-centric methods of (3.2) for reduction of multivariate dimensionality. While Longley used seven variables in his ill-conditioned regression model, here we use YEAR = 1947 to 1962 only as labels for points, rather than as an exact "linear trend" variable. The six US economic

variables used here are X_1 = GNP deflator, X_2 = GNP, X_3 = number Unemployed, X_4 = size of Armed Forces, X_5 = total Population and X_6 = total Employed.

This example is rather typical of cases I have explored, especially when N greatly exceeds p. Specifically, the optimal affine reduction in Mahalanobis dimensionality depicted in Figure 4 has a "Dyson swarm" interpretation, featuring a characteristic "hole" surrounding the explicit origin.

7 Discussion

In the terminology of Glaeser(2006), "Researcher Initiatives" can deliberately bias findings towards the perspective of one particular subset of stakeholders. For example, it is well know that including several different quantitative X-variables that are surrogate measures of a single construct in a PCA tends to orient the first principal axis to represent that construct. The corresponding phenomena in an origin-centric analysis would be to include several points that are actually coincident, thereby placing greater emphasis upon recovering the distance of this composite from all other points within the scatter.

Origin-centric methods allow the scatter to be visualized much like a "Dyson Swarm" for a finite universe of N or fewer points. For example, a q=3 dimensional recovered scatter might be primarily viewed as a 2-dimensional small-scale conformal mapping where radial distance information is disregarded. Alternatively, the size of each of the N plotting symbols displayed on this sort of "sky view" map could represent some monotone function of its recovered radius.

Nonnegative Matrix Factorization (NMF) techniques rely upon locating the origin so that all N initial points have p strictly nonnegative coordinates. In other words, all initial points must lie on or within the boundaries of the "first" Cartesian 2^p-ant of p-dimensional Euclidean space. Similarly, the q dimensions to be recovered using NMF are to be represented as vectors emanating from the origin that have strictly nonnegative direction cosines because they point exclusively into this same first Cartesian 2^p-ant.

8 Appendix: Affine Invariance and Commutativity

Linear algebra is typically used to manipulate data structures, consisting of vectors and matrices of real numbers, in multivariate statistical analyses. Insights into the analytical geometry of such manipulations is provided by visualizing the data as consisting of a scatter of N points within a p-dimensional Euclidean space.

A fundamental property of Euclidean space is that distance is computed from coordinates relative to orthogonal axes using the Pythagorean theorem.

A N × p matrix of real numbers, $\mathbf{Y}$, is said to represent a nonsingular affine transformation of an initial N × p matrix of real numbers, $\mathbf{X}$, if and only if there exists a 1 × p "translation" vector $\mathbf{a}'$ and a nonsingular p × p matrix $\mathbf{B}$ of real values such that

$$\mathbf{Y} = \mathbf{1}\mathbf{a}' + \mathbf{X}\mathbf{B}. \tag{8.1}$$

In particular, if any row of $\mathbf{X}$ consists of all zeros, the corresponding row of $\mathbf{Y}$ will be $\mathbf{a}'$. The notation of equation (8.1) differs slightly from that of Obenchain(1971) simply because the $\mathbf{X}$ and $\mathbf{Y}$ matrices have been transposed here.

A statistic $S(Y)$, consisting of a scalar, vector or matrix of real numbers, will be said to be affine invariant if and only if its numerical value when computed after any nonsingular affine transformation has been applied to a dataset is identical to its value before such transformation:

$$S(\mathbf{Y}) = S(\mathbf{X}) \text{ ...for every translation } \mathbf{a} \text{ and nonsingular } \mathbf{B} \tag{8.2}$$

The $\mathbf{H}$ matrix of (3.1) is not uniquely determined. For example, any of the r columns of $\mathbf{H}$ (each N × 1) can always be multiplied by -1 if the corresponding row of $\mathbf{G}'$ (1 × p) is also multiplied by -1. On the other hand, $\mathbf{H}\mathbf{H}'$ is an idempotent and symmetric (N × N) matrix that is uniquely determined and corresponds to the orthogonal projection matrix (linear operator) that characterizes the column space of the $(\mathbf{I} - \mathbf{1}\gamma')\mathbf{X}$ matrix; see Rao(1973), pages 46-48. In fact, Obenchain(1971) showed that $\mathbf{H}\mathbf{H}'$ and $\mathbf{D}^{(2)}$ are equivalent "maximal affine invariant" statistics which can be viewed as a standardized, canonical form for a given (N × p) data matrix of "centered" scalar values subject to nonsingular linear $\mathbf{B}$ transformations, (8.1).

The $\mathbf{H}\mathbf{H}'$ and $\mathbf{D}^{(2)}$ matrices are unique but distinct, and each can be computed from the other. PCA focuses upon reproducing $(\mathbf{I} - \mathbf{1}\gamma')\mathbf{X}\mathbf{X}'(\mathbf{I} - \gamma\mathbf{1}')$ by decomposing it into ordered and additive components. PCA fails to reduce dimensionality when applied to $\mathbf{H}\mathbf{H}'$ because all of its components are equally good or bad and, thus, cannot be uniquely ordered.

By focusing upon reproducing $\mathbf{D}^{(2)}$ instead of $\mathbf{H}\mathbf{H}'$, origin-centric methods can reduce the dimensionality of configurations that PCA cannot. There appear to be at least two separate parts to the price-one-has-to-pay to achieve affine invariance properties that ignore vagaries caused by analyst choice of variable scalings and/or confounding between variables. First of all, recovered configurations will not be additive; the optimal solution of dimension (q+1) usually does not exactly "contain" the optimal solution of dimension q. Secondly, minimizing $\mathbf{Norm}^2$ of equation (2.3) is a problem of constrained optimization subject to multiple local minima. Fortunately, modern numerical search software, such as the "optimx" R-package of Nash and Varadhan(2011), is now available to researchers.

9 Appendix: Detail on Data Centering and Affine Median Vectors

A N $\times$ p matrix of real numbers, $\mathbf{Y}$, will be said to be a "centered" version of a N $\times$ p matrix of real numbers, $\mathbf{X}$, if and only if

$$\mathbf{Y} = (\mathbf{I} - \mathbf{1}\gamma')\mathbf{X}, \tag{9.1}$$

where the $1 \times N$ row vector γ' is any generalized inverse of the N $\times$ 1 column vector of all ones, $\mathbf{1}$. In other words, any $1 \times N$ row vector of real values that sum to 1, $\gamma'1 = 1$, is a valid choice for γ' in (9.1).

The "centering" operation of equation (9.1) frequently yields $\mathbf{Y}$ coordinates that differ from those of the initial $\mathbf{X}$ matrix. In these cases, the location of the implicit origin, $\mathbf{0}$, has been shifted via a simple "translation" of $\mathbf{Y}$ coordinates.

The unique Moore-Penrose inverse of $\mathbf{1}$, usually denoted by $\mathbf{1}^+$, is the row vector with each of its N entries equal to $1/N$. This choice for γ' corresponds to centering the $\mathbf{X}$ matrix at its overall mean vector, i.e. at its traditional centroid.

When γ_1 and γ_2 are two possibly different choices for the γ vector of (9.1), the following equality is easily verified by direct multiplication and algebraic simplification:

$$(\mathbf{I} - \mathbf{1}\gamma_2')(\mathbf{I} - \mathbf{1}\gamma_1') = \mathbf{I} - \mathbf{1}\gamma_2'. \tag{9.2}$$

When $\gamma_1 = \gamma_2$, equation (9.2) shows that all valid choices for the γ vector of (9.1) yield a centering matrix, $\mathbf{I} - \mathbf{1}\gamma'$, that is idempotent and, thus, corresponds to a geometrical projection in N space that generally is oblique. In fact, the choice $\gamma_1' = \gamma_2' = \mathbf{1}^+$ is the only choice that makes the $\mathbf{I} - \mathbf{1}\mathbf{1}^+$ matrix symmetric as well as idempotent, thus corresponding to a strictly orthogonal projection in N space.

Perhaps, an even more interesting result follows from equation (9.2) when γ_1 and γ_2 are distinct generalized inverses. In these cases, note that any valid, final choice of centering, as in equation (9.1), will simply "wipe-out" and "replace" any and all previously applied choices of centering of form (9.1). The corresponding implication for (3.1) is that changing one's choice of γ vector typically changes the $\mathbf{H}$ matrix so that it becomes orthogonal to the new γ.

To center an $\mathbf{X}$ matrix at the point corresponding to its i-th row, the i-th element of γ would then be a 1 and all other elements would be null.

A multivariate median vector can be defined, as in Obenchain(1972a), using "convex hull" concepts to assure that the γ vector is an invariant function under all strictly nonsingular affine transformations of a given $\mathbf{X}$ scatter, i.e. where no systematic reductions in dimensionality are being enforced. In direct analogy with the concept of a univariate median, successive "exterior" convex hulls (which are preserved under nonsingular affine transformations) of the scatter initially spanning p-dimensional Euclidean space would be determined and successively "peeled away." Whenever two or more points on an exterior hull are coincident, only one such point is set aside

(assigned a γ weight of zero) at each stage of such peeling. The final convex hull remaining at the end of this sequence will contain neither any coincident points nor any strictly interior points. The affine invariant γ vector, which defines the corresponding affine commutative median vector, $\gamma'\mathbf{X} = -\mathbf{a}'$ of equation (8.1), gives equal, positive weight to each of the unique points on the innermost hull and zero weight to all other points.

10 Appendix: Restrictions to Distinct Points, possibly Weighted.

In this penultimate appendix, we comment on the role of mutually coincident points in determining the dimensionality of a given scatter.

Since coincident points clearly cannot increase dimensionality and always remain coincident in the affine reduction formulation of (3.2), it makes good sense computationally to limit attention to scatters of only distinct, non-coincident points. Since no reasonable method will fail to exactly reproduce the zero diagonal elements of the $\mathbf{D}^{(2)}$ matrix of (2.2), minimizing the squared norm of equation (2.3) rightly focuses on minimizing the sum of the $N(N-1)/2$ terms with $1 \leq i < j \leq N$, where N now denotes the total number of distinct points. On the other hand, individual terms in this summation could certainly be differently "weighted" to account for the presence or absence of additional points coincident with either the i-th or the j-th distinct point.

Limiting attention to unique points and affine reduction methods also assures, when many variables are analyzed for a fixed number of points, that the maximum possible value of r will indeed equal $(N-1) \leq p$ in equation (3.1). For an arbitrary configuration of N unique points, one then knows in advance that $\mathbf{H}'\mathbf{H}$ will ultimately become the $(N-1) \times (N-1)$ identity matrix as the number of columns of $\mathbf{X}$ is arbitrarily increased. Furthermore, in this same limit, the corresponding $\mathbf{HH}'$ matrix will also approach the $\mathbf{I} - \gamma\gamma^+$ matrix, which is the N × N orthogonal projection matrix for the vector space orthogonal to γ.

11 Appendix: Affine Invariance and the Restriction to p less than or equal to $N-1$.

When $\gamma' = \mathbf{1}^+$, the large-p asymptotic H-configuration is that of N points equally spaced in $N-1$ dimensions, which corresponds to a sparse, high-dimensional Dyson sphere with no point coincident with the origin. Specifically, the squared distance between any 2 of these N points will be 2, and the squared distance from each to the origin will be $d_0^2 = (N-1)/N$.

When all of the elements of γ except one are zeros, the asymptotic configuration will depict that single point as being coincident with the origin while the remaining $N-1$ points are again equally spaced. While the squared distance between any two of these $N-1$ points will still be 2, the squared distance from each of them to the origin point will now be $d_0^2 = 1$. The first $N-1$ rows of the asymptotic $\mathbf{H}$ matrix may then be any $(N-1) \times (N-1)$ orthogonal matrix (such as some permutation of the rows and columns of the $\mathbf{I}$ matrix), while the final row is null. This configuration again spans $N-1$ dimensions and depicts another sparse, high-dimensional Dyson sphere as well as its origin point.

12   Acknowledgments

13   References

Dyson, FJ. Search for Artificial Stellar Sources of Infra-Red Radiation. *Science*, 1960, 131, 1667-1668.

Gabriel, KR. The Biplot Graphic Display of Matrices with Application to Principal Component Analysis. *Biometrika*, 1971, 58(3), 453-467.

Gabriel, KR. Analysis of Meteorological Data by Means of Canonical Decomposition and Biplots. *Journal of Applied Meteorology*, 1972, 11, 1071-1077.

Glaeser, EL. Researcher incentives and empirical methods. *Harvard Institute of Economic Research, discussion paper #2122*. http://scholar.harvard.edu/glaeser 2006.

Gower, JC. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 1966, 53, 325-338.

Gower, JC. Adding a point to vector diagrams in multivariate analysis. *Biometrika*, 1968, 55, 582-585.

Gower, JC. Euclidean distance geometry, *Math. Sci.*, 1982, 1, 1-14.

Gower, JC. Properties of Euclidean and Non-Euclidean Distance Matrices *Linear Algebra and its Applications*, 1985, 67, 81-97.

Lee, DD and Seung, HS. Learning the parts of objects by non-negative matrix factorization. *Nature*, 1999, 401(6755), 788-791.

Longley, JW. An appraisal of least squares programs for the electronic computer from the point of view of the user. *Journal of the American Statistical Association*, 1967, 62, 819-841.

Nash, JC and Varadhan, R. Unifying Optimization Algorithms to Aid Software System Users: optimx for R. *Journal of Statistical Software*, 2011, 43(9), 1-14.

Mahalanobis, PC. On the generalized distance in statistics. *Proc. Nat. Inst. Sci. India, Part A.* 1935, 12, 49-55.

Obenchain, RL. Multivariate procedures invariant under linear transformations. *Annals of Mathematical Statistics*, 1971, 42, 1569-1578.

Obenchain, RL. Correction Note: Multivariate procedures invariant under linear transformations. *Annals of Mathematical Statistics*, 1972a, 43, 1742-1743.

Obenchain, RL. Regression optimality of principal components. *Annals of Mathematical Statistics*, 1972b, 43, 1317-1319.

Obenchain, RL. Affine Multidimensional Scaling. Presentation at *Third International Symposium on Multivariate Analysis*, Dayton, OH, 1972c.

Obenchain, RL. The geometry of linear reparameterizations in regression. *Contributions to Statistics: Essays in Honor of Norman L. Johnson*, ed. P. K. Sen, Amsterdam: North-Holland 1983, 347-365.

Obenchain, RL. ICE Preference Maps: Nonlinear Generalizations of Net Benefit and Acceptability. *Health Services and Outcomes Research Methodology*, SpringerLink (open access), 2008, 8, 31-56. (two-dimensional, origin-centric maps of economic preferences.)

Okamoto, M. Optimality of principal components. *Multivariate Analysis, 2* P. R. Krishnaiah, ed. New York: Academic Press, 1969, 673-685.

Rao, CR. Linear Statistical Inference and Its Applications, 2nd Edition. New York: John Wiley and Sons, 1973.

Schoenberg, IJ. Remarks to Maurice Freehet's article "Sur la definition axiomatique d'une classe d'espaces vectoriels distancies applicables vectoriellement sur l'espace de Hilbert", *Ann. Math.* 1935, 36, 724-732.

Torgerson, WS. *Theory and methods of scaling.* New York: Wiley, 1958.

Young, G. and Householder, AS. Discussion of a set of points in terms of their mutual distances. *Psychometrika*, 1938, 3, 19-22.
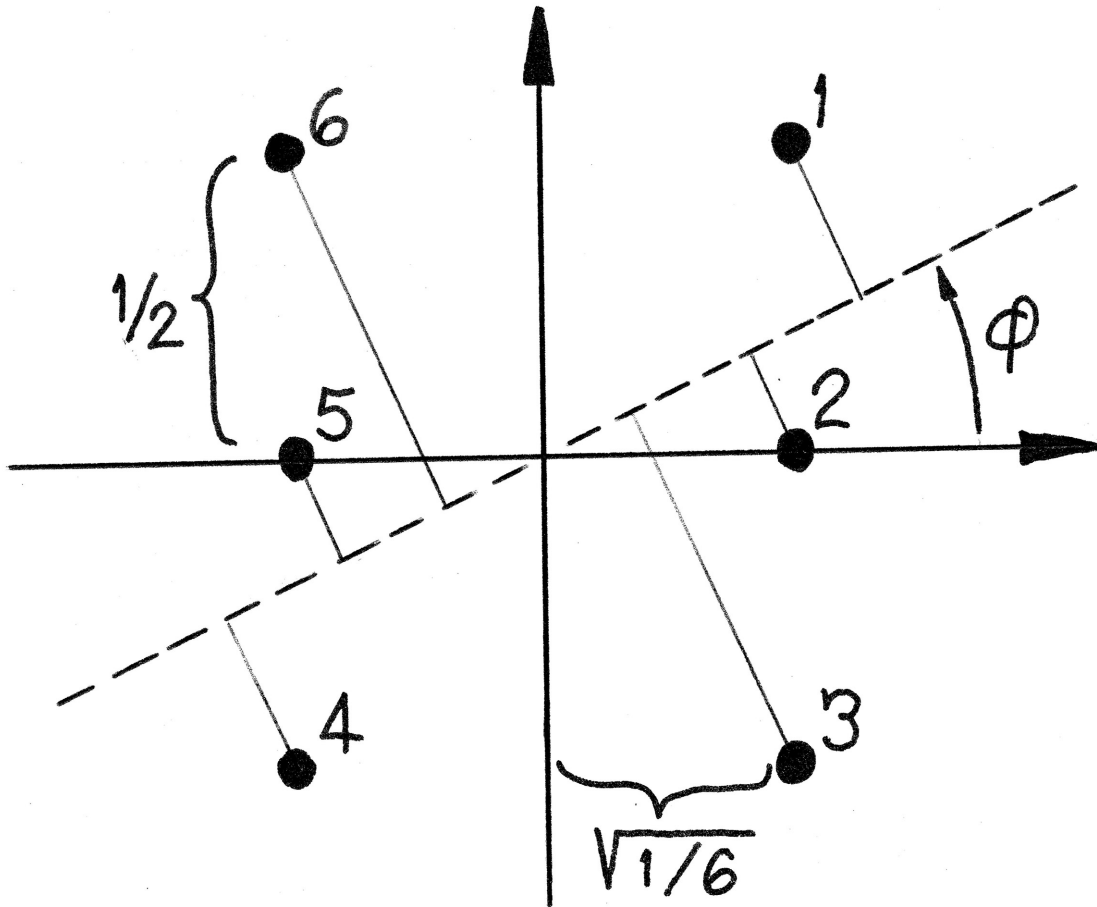
Figure 1: Mahalanobis H Configuration for 6 Points in 2 Dimensions. Numerical searches for the best q=1 representation can use, as their starting points, orthogonal projections of the 6 points onto the line through the origin making angle ϕ with the horizontal axis. Initial values for the **B** matrix of (3.2) are then simply 2-element column vectors of the form $(sin(\phi), cos(\phi))'$, and the objective of the search is to find the best length for **B** in each given direction.
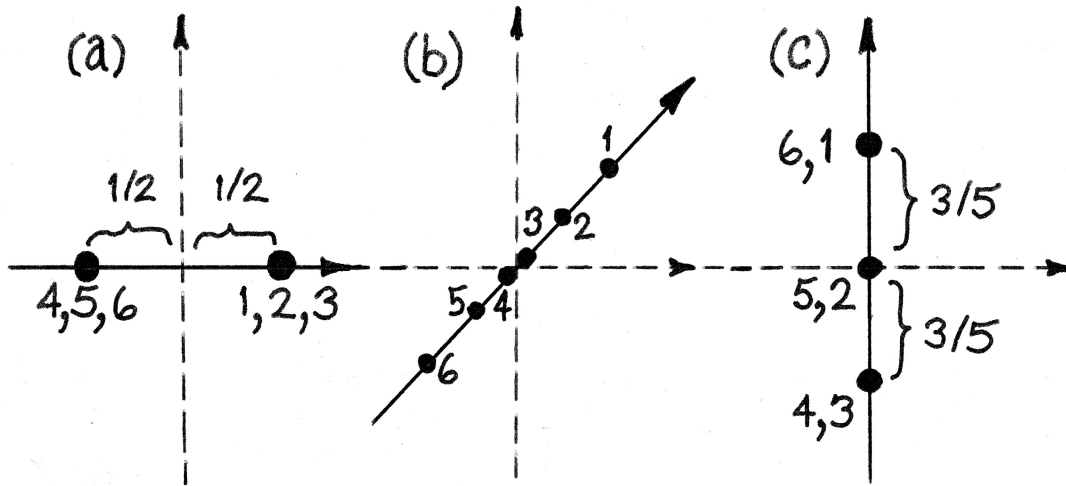
Figure 2: Depiction of three Local Minima of **Norm**^{2} for Dimension q=1. While a pair of local minima occur in case (a) at 3.50 and in case (b) at 4.80, the global minimum of 3.11 corresponds to case (c).
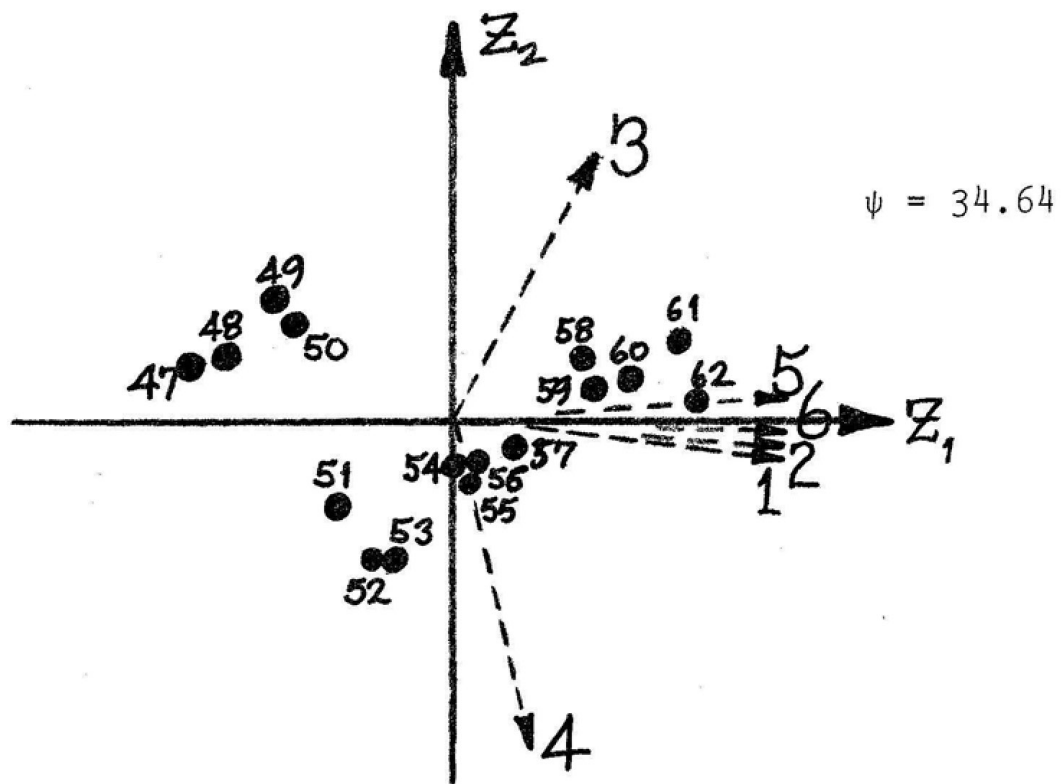
Figure 3: The PCA solution of Dimension q=2 for the Longley data when standardized to "correlation" form. It is clear from the labels on the points as well as from the direction cosines for Axes 1, 2, 5 and 6 that this recovered configuration is best interpreted as representing "linear growth" over time.
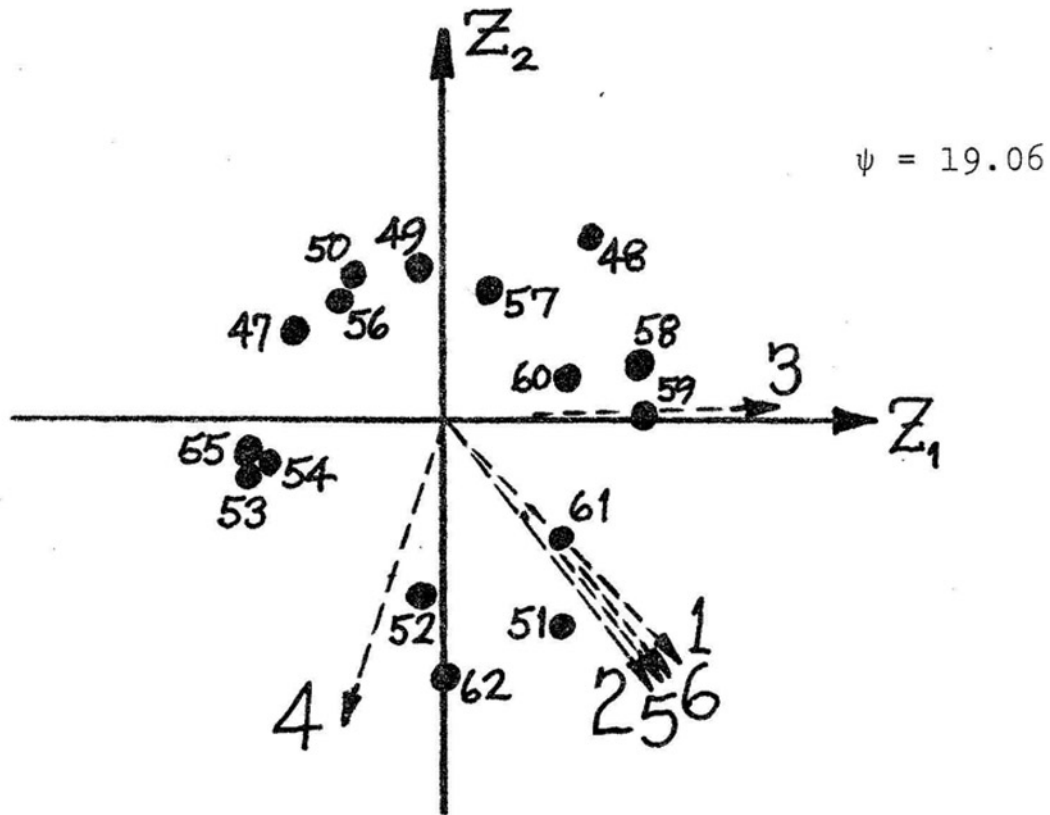
Figure 4: This is the optimal origin-centric representation in q=2 dimensions for the Longley data when standardized to Mahalanobis H form. The "year" labels on points suggest that time variation is being depicted here as clockwise or counter-clockwise rotations. In particular, the direction cosines for Axes 3 and 4 suggest that these two measures are the true, primary determinants of year-to-year variation in q=2 dimensions.